The Network Solution For Electron Identification in a Wide Momentum Region with a TRD

N. Kuropatkin*,† and R. Zukanovich Funchal‡

Instituto de Física da Universidade de São Paulo
05389-970 C.P. 66318 – São Paulo, SP
Brazil

A Feed Forward Error Back Propagation Artificial Neural Network (ANN) algorithm is developed for electron/positron identification in a wide momentum region (10 - 300 GeV/c). The method was proposed for the Transition Radiation Detector of the E781 experiment at Fermilab. The package consists of two parts:
- the program for the ANN training;
- the particle classification subroutine.

Both parts are built using the object oriented technique and C++ language. The particle identification algorithm is wrapped in FORTRAN closers to be used in the E781 off-line program. The package performance was tested in comparison with the likelihood ratio method using Monte Carlo generated data. Our study has demonstrated the excellent ability of the ANN to learn even small details of the detector response function. The ANN solution gives the same performance and behavior as the likelihood method when using Monte Carlo data with known detector parameters. It demonstrates that if trained with experimental data the package can provide a very good solution to the classification problem of e^+/e^- tracks.

INTRODUCTION

In this work we have studied the ability of an Artificial Neural Network (ANN) based method applied to a Transition Radiation Detector (TRD) to select e^+/e^- tracks in a high energy fixed target experiment.

In hadroproduction fixed target experiments, such as the E781 experiment at Fermilab [1], the primary interaction produces many tracks in a narrow solid angle around the beam direction. Most of these particles will not be of leptonic nature. Nevertheless they can be responsible directly or indirectly (if for instance they suffer secondary interactions) for the appearance of ionization clusters in the TRD which can fake the e^+/e^- signal under investigation.

We have used GE781 [2], the GEANT [3] based Monte Carlo simulation of the E781 apparatus in our studies. The TRD model we are working with is made of 6 blocks, similar to the ones described in Ref. [4], each one composed by a radiator (220 foils of CH_2) and a Xe-CH_4 filled proportional chamber with the ability to measure the x-coordinate of the track. The detector operates in the cluster counting mode [5]. Ionization clusters were counted only along each track direction to reduce the influence of the background.

The momentum region of the tracks under investigation is from 10 to 300 GeV/c, which corresponds to the growing part of the Transition Radiation yield curve for pions [5]. Above this momentum it is more difficult to distinguish pions from electrons using this detector.

To achieve appropriate background suppression as well as good signal efficiency we have exploited the $\gamma = E/m$ (where E is the particle energy and m is its mass) factor dependence of the detector response in our algorithm.

In the following we describe and discuss the maximum likelihood ratio technique, our ANN approach and compare their performance with Monte Carlo simulated data.

---

*On leave of absence from Petersburg Nuclear Physics Institute,Gatchina, Russia.

†E-mail: kuropat@charme.if.usp.br.

‡E-mail: zukanov@charme.if.usp.br.

DESCRIPTION OF THE TRD LIKELIHOOD METHOD

This method was proposed for TRD detectors by M. L. Cherry et al. [6] and was demonstrated to give better performance than traditional methods in Ref. [7–9].

The likelihood function is built in a way to classify particles into two categories : e^+/e^- (type 1) and others, mostly pions, (type 2). So if a particle of type $i = 1, 2$ with Lorenz factor γ_i generates a sample $X=\{x_1, x_2, ..., x_6\}$ of transition radiation clusters along its track we can define the probability $P(X|i, \gamma_i)$ of this event as

$$P(X|i, \gamma_i) = \prod_{k=1}^{6} P(x_k|\gamma_i). \tag{1}$$

The probability density functions $P(x_k|\gamma_i)$ were calculated using the detector response function.

We can define the likelihood ratio for particle of type i as

$$\mathcal{L}_i(X|\gamma_i) = \frac{P(X|i, \gamma_i)}{\sum_{j=1}^{2} P(X|j, \gamma_j)}, \tag{2}$$

which is restricted to the interval $0 \le \mathcal{L}_i \le 1$, and use this ratio as an indicator of the particle type. For each track one can calculate the above ratio in two possible hypothesis. We expect that this ratio will be closer to 1 whenever the hypothesis is correct and closer to zero whenever it is wrong.

DESCRIPTION OF THE TRD NEURAL NETWORK

In order to achieve the maximal performance in the electron identification with TRD we have developed an algorithm using a Feed Forward Error Back Propagation Artificial Neural Network [10]. A similar application was developed in Ref. [11] for 10 modules of TRD used for cosmic ray lepton identification.

As we have 6 modules of TRD we will use 6 input nodes with linear response function that will receive the cluster sum along the track in each TRD block normalized to unity. To explicitly take into account the γ dependence of the detector response and consequently increase the momentum region in which the algorithm can provide an efficient classification of tracks, an extra linear node was introduced. This node was fed with a normalized to one γ factor calculated in the pion hypothesis, i.e.

$$\text{node 7 activation} = \frac{E}{m_\pi \gamma_{\text{cut}}}, \tag{3}$$

where E is the energy of the particle, m_π the mass of the pion and γ_{cut} was chosen to be 2200, which corresponds to pions of about 310 GeV/c. For γ greater than γ_{cut} the node 7 activation is set equal to one.

We start with 7 input nodes and would like to have a similar classification for particles here as in the previous method, that is, two output nodes. So according to Kolmogorov theorem [12] we can approximate our classification function with 15 nodes in the hidden layer. This defines the structure of the network as 7+15+2 nodes, as presented in Fig. 1.
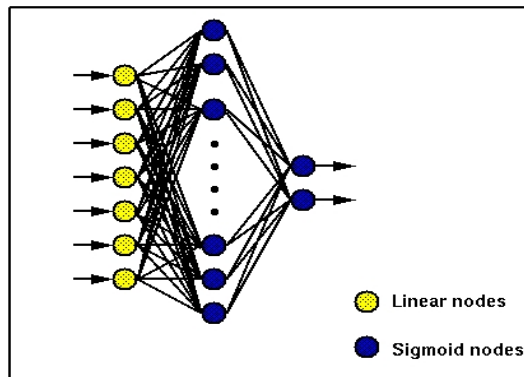

FIG. 1. The TRD neural network structure.

The first layer of nodes are fully connected to the second layer of 15 nodes with a sigmoid response function which is followed by the output layer of 2 nodes (res 1 and res 2). Each neuron of one layer receives as input the outputs of all neurons from the previous layer with weights defined by the synaptic matrix W. The activation level of the output nodes will provide the track classification.

The sigmoid function we have used is defined as following:

$$F(x, b) = \frac{1}{1 + \exp(-a(x - b))}, \tag{4}$$

where b is the neuron threshold, and a is a gain. It is clear that the responses of the two output nodes are bound to the interval $[0, 1]$ and provide a similar classification scheme as in the previous method.

The training process was performed with the standard back-propagation technique [10] during which the corrections to the synaptic matrix elements W_{lk} were calculated according to the rule :

$$\Delta W_{lk}(i) = -S \frac{\partial E[W]}{\partial W_{lk}} + M \Delta W_{lk}(i - 1), \tag{5}$$

where $\Delta W_{lk}(i)$ is the correction to the synaptic matrix elements W_{lk} after i steps, E[W] is the summed square error function, S is the learning rate parameter and $M \Delta W_{lk}(i-1)$ is the momentum term used to avoid sudden oscillations. We have used in our implementation $S = 0.1$ and $M = 0.3$.

The method consists of two independent parts:

- the training program;

- the particle classification subroutine.

Both parts were written in C++ using the Object Oriented (OO) approach. As the basis for our algorithms we have used modified classes provided by Robert Klapper [13]. The training procedure can use either Monte Carlo or experimental data. The result of the training program is a file describing the ANN structure and parameters. This file is used then by the particle identification algorithm in the initialization stage, this permits to dynamically update the ANN parameters. The particle identification algorithm is wrapped in FORTRAN closers to be used in the E781 off-line program.

To train our network we have used Monte Carlo simulated data with plane momentum distribution in the whole momentum region. For this the detector response function was studied and implemented in the Monte Carlo. It is in principle also possible to use experimental data for this purpose, but in this case we need some independent tagging of electrons to prepare pure samples of particles with high statistics and in a wide momentum region.

COMPARISON OF THE METHODS

To compare the performance of these two methods we have used single tracks generated in the whole momentum region and Monte Carlo simulated hadronic interactions enriched with leptonic processes. In the last case we have exactly the same momentum distribution of particles as expected in the experiment.

To choose a suitable efficiency to contamination ratio working point we can apply correlated cuts in the feature space of $\mathcal{L}_e$ (in the e^+/e^- hypothesis) and $\mathcal{L}_h$ (in the other particles hypothesis) for the likelihood method and in the feature space of res 1 and res 2 for the artificial neural network. This can be done in the following way:

$$\mathcal{L}_e > \text{cut} \quad \text{and} \quad \mathcal{L}_h \leq 1. - \text{cut}, \tag{6}$$

for the likelihood ratio case and

$$\text{res } 1 > \text{cut} \quad \text{and} \quad \text{res } 2 \leq 1. - \text{cut}, \tag{7}$$

for the network, where cut is any real value from zero to one. Changing the value of the cut we can build plots of the hadronic contamination as a function of electron detection efficiency for these two methods.
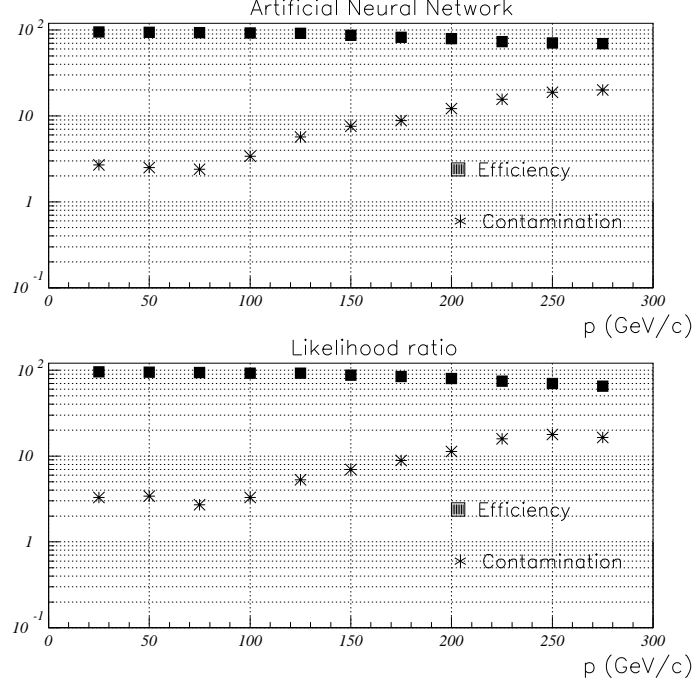
3

FIG. 2. Efficiency and contamination for the two algorithms as a function of particle momenta.

We will start our comparison of the methods acceptance and rejection power using single tracks generated in a wide momentum range from 10 to 300 GeV/c. The momentum region was subdivided in bins of 30 GeV/c, and the corresponding parameters were calculated for each bin. To compare the methods we have selected cuts in such a way to have approximately equal efficiencies for both methods in the first momentum bin.

As one can see in Fig. 2 both methods show similar behavior and the revealed growth of the contamination and decrease of the efficiency corresponds to the degradation with momentum growth of the classification power of the detector itself .
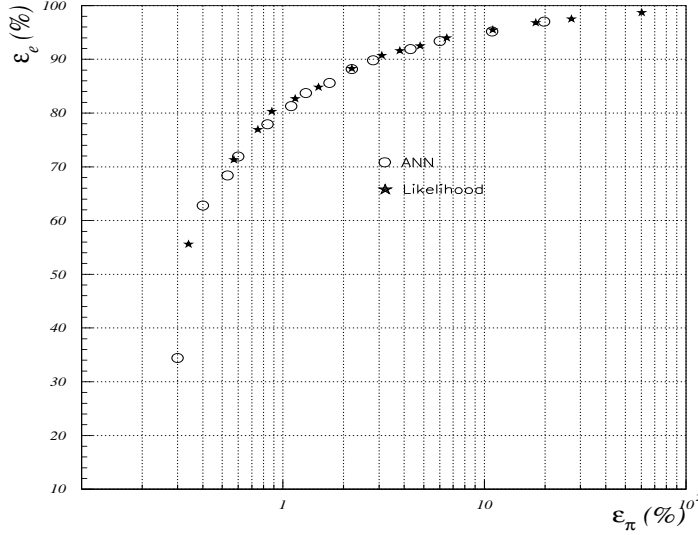
FIG. 3. Electron identification efficiency (ϵ_e) versus hadronic contamination (ϵ_π), taken from simulation of hadroproduction data for the two methods under study.

Simulation of the hadronic interaction data

To compare the two methods in experimental conditions, including hadronic background, secondary interactions and real momentum distribution we decided to use Monte Carlo generated hadronic interactions with complete simulation of the detector by the GE781 package. The hadronic background, as well as the electromagnetic one, simulated by the package is supposed to be very close to the experimental one.

The e^+/e^- efficiency versus the contamination by all other particles for such sample is shown in Fig. 3. One can see that the two methods are indistinguishable. We really can not give preference to any of these two methods from the performance point of view.

CONCLUSIONS

We have studied the performance of the ANN solution to particle classification with a TRD which directly use the momentum information of the tracks.

The ANN solution gives the same performance and behavior as the likelihood method when using Monte Carlo data with known detector parameters. It demonstrates that if trained with experimental data the package can provide a very good solution if not the best one as the ANN can learn unknown properties of the detector and of the experimental conditions that can not be implemented in the likelihood method.

ACKNOWLEDGMENTS

---

[1] FNAL E781 (SELEX) proposal (unpublished), July 1993; J. Russ, "SELEX – Hadroproduction of Charm Baryons out to Large x_F", presented at "Production and decay of Hyperon, Charm and Beauty Hadrons", Strasbourg, France (1995).

[2] G. Davidenko *et al.*, GE781: a Monte Carlo package for fixed target experiments, in: R. Shellard and T. D. Nguyen, eds., *Proceedings of the International Conference on Computing in High Energy Physics'95* (World Scientific, Singapore, 1996) p. 832.

[3] R. Brun, F. Bruyant, M. Marie, A. C. McPherson, and P. Zanarini, GEANT3, DD/EE/81-1, CERN (1987); GEANT Detector Description and Simulation Tool, Application Software Group, CND, CERN Program Library Long Writeup W5013 (1994).

[4] V. T. Grachev *et al.*, in *Proceedings of the Symposium on Particle Identification at High Luminosity Hadron Colliders*, edited by T. J. Gourlay and J. G. Morfin (Fermi National Accelerator Laboratory,Batavia, Il. 1989) p. 415; V. A. Andreev *et al.*, preprint LENINGRAD-86-1186 (1986).

[5] B. Dolgoshein, Nucl. Instr. and Meth. A326 (1993) 434.

[6] M. L. Cherry *et al.*, Nucl. Instr. and Meth. 115 (1974) 141.

[7] A. Büngener *et al.*, Nucl. Instr. and Meth. 214 (1983) 261.

[8] K.-K. Tang, The Astrophysical Journal, 278 (1984), 881.

[9] R. D. Appuhn *et al.*, Nucl. Instr. and Meth. A263 (1988) 309.

[10] C. Peterson and T. Rögnvaldson, "An Introduction to Artificial Neural Networks", LUTP 91-23, Lectures given at the 1991 CERN School of Computing, Ystad (1991).

[11] R. Bellotti *et al.*, Computer Phys. Commun. 78 (1993) 17.

[12] A. N. Kolmogorov, Dokl. Akad. Nauk. USSR 114 (1957) 953.

[13] R. Klapper http://www.cs.nyu.edu∼klap7794/spiderweb2.html.